COMPUTER RESEARCH AND MODELING
2026 VOL. 18 NO. 1 P. 185–200
DOI: 10.20537/2076-7633-2026-18-1-185-200

Ки&М

ANALYSIS AND MODELING OF COMPLEX LIVING SYSTEMS

UDC: 004.85:614.4

Reducing computational complexity in agent-based epidemiological model calibration: application of deep learning surrogates

A. Darwish[1,a], V. N. Leonenko[1,2,b]

[1]ITMO University,
49 Kronverksky pr., St. Petersburg, 197101, Russia
[2]Smorodintsev Research Institute for Influenza,
15/17 Professora Popova st., St. Petersburg, 197022, Russia

E-mail: [a] adarwish@itmo.ru, [b] vnleonenko@itmo.ru

Received 20.10.2025, after completion — 10.12.2025.
Accepted for publication 12.12.2025.

Acute respiratory infections are a major public health concern because they are the leading cause of illness and death in many countries. Therefore, there is great interest in developing models and methods capable of modeling the spread of these infections within communities, with the aim of controlling outbreaks and preventing their spread. Agent-based models (ABM) are one of the most important tools in epidemiological research for modeling epidemic dynamics in realistic populations, but they face significant challenges in terms of computational complexity in their operation and calibration of epidemiological data, as parameter estimation typically requires repeated simulations across large parameter spaces to determine plausible values for key epidemiological parameters. This paper addresses the problem of alleviating computational constraints in the inverse problem of calibrating an ABM model for simulating the spread of respiratory infections in Saint Petersburg. The paper proposes the application of machine learning surrogate to link epidemic trajectories to underlying epidemiological parameters, enabling them to quickly infer parameter estimates from observed epidemic data. This is done by formulating the task of calibrating ABMs against epidemiological data as a supervised learning problem, where sequences extracted from epidemiological trajectories are associated with underlying epidemiological parameters. The research was based on evaluating the performance of attention-based sequence modeling, probabilistic deep learning, and distributional regression for inferring parameter estimates from truncated sequences of epidemic trajectories. Experimental evaluations have demonstrated the effectiveness of this approach and its practical and straightforward application. The results also indicated the superiority of attention-based sequence modeling, as it showed more consistent performance across metrics and horizons in accurate parameter estimation and credible uncertainty quantification. Distributional regression modeling also showed good performance with specific strengths in point accuracy while probabilistic deep learning performed poorly, especially at longer input horizons.

Keywords: epidemiology, ABM simulations, machine learning (ML), inverse problems, curse of dimensions

Citation: *Computer Research and Modeling*, 2026, vol. 18, no. 1, pp. 185–200.

This work was supported by the Russian Science Foundation in accordance with Agreement No. 24-25-20104 dated 12.04.2024 and the St. Petersburg State Autonomous Institution “Fund for the Support of Scientific, Scientific-Technical, Innovative Activities” on the provision of a regional grant for co-financing the project under agreement 24-25-20104 dated 22.05.2024 dated 03.06.2024 No. 169.

**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2026 Т. 18 № 1 С. 185–200**
DOI: 10.20537/2076-7633-2026-18-1-185-200

Ки&М

УДК: 004.85:614.4

Снижение вычислительной сложности при калибровке агентных эпидемиологических моделей: применение суррогатных моделей глубокого обучения

А. Дарвиш[1,a], В. Н. Леоненко[1,2,b]

[1]Университет ИТМО,
Россия, 197101, г. Санкт-Петербург, Кронверкский пр., д. 49
[2]ФГБУ «НИИ гриппа им. А. А. Смородинцева» Минздрава России,
Россия, 197022, ул. г. Санкт-Петербург, ул. Профессора Попова, д. 15/17

E-mail: [a] adarwish@itmo.ru, [b] vnleonenko@itmo.ru

*Получено 20.10.2025, после доработки — 10.12.2025.
Принято к публикации 12.12.2025.*

Острые респираторные инфекции представляют собой серьезную проблему общественного здравоохранения, поскольку являются одной из причин заболеваемости и смерти во многих странах. В связи с этим существует большой интерес к разработке моделей и методов, позволяющих моделировать распространение этих инфекций в сообществах с целью контроля вспышек и предотвращения их распространения. Агентные модели (АМ) являются одним из важнейших инструментов эпидемиологических исследований для моделирования динамики эпидемий в реальных популяциях, но они сталкиваются со значительными трудностями, связанными с вычислительной сложностью при их использовании и калибровке эпидемиологических данных, поскольку оценка параметров обычно требует многократного моделирования в больших пространствах параметров для определения правдоподобных значений ключевых эпидемиологических показателей. В данной статье рассматривается проблема снижения вычислительных ограничений в обратной задаче калибровки АМ для моделирования распространения респираторных инфекций в Санкт-Петербурге. В статье предлагается применение суррогатного машинного обучения для связи траекторий эпидемий с лежащими в их основе эпидемиологическими параметрами, что позволяет быстро выводить оценки параметров на основе наблюдаемых эпидемических данных. Это достигается путем формулировки задачи калибровки АМ по эпидемиологическим данным как задачи контролируемого обучения, в которой последовательности, извлеченные из эпидемиологических траекторий, связываются с базовыми эпидемиологическими параметрами. Исследование было основано на оценке эффективности моделирования последовательностей на основе внимания, вероятностного глубокого обучения и распределительной регрессии для вывода оценок параметров из усеченных последовательностей эпидемических траекторий. Экспериментальные оценки продемонстрировали эффективность данного подхода и его практическое и простое применение. Результаты также указали на превосходство моделирования последовательностей на основе внимания, поскольку оно показало более стабильную производительность по всем метрикам и горизонтам, обеспечивая точную оценку параметров и достоверное количественное определение неопределенности. Моделирование распределительной регрессии также показало хорошую производительность, особенно с точки зрения точности точек, в то время как вероятностное глубокое обучение показало плохую производительность, особенно при более длительных входных горизонтах.

Ключевые слова: эпидемиология, агентное моделирование, машинное обучение (ML), обратные задачи, проклятие размерности

Работа выполнена при финансовой поддержке Российского научного фонда в соответствии с Соглашением № 24-25-20104 от 12.04.2024 г. и Санкт-Петербургского государственного автономного учреждения «Фонд поддержки научной, научно-технической, инновационной деятельности» о предоставлении регионального гранта для софинансирования проекта в рамках соглашения № 24-25-20104 от 22.05.2024 г. от 03.06.2024 г. № 169.

Introduction

Background

Acute respiratory infections in general, and viral infections such as influenza and COVID-19 in particular, are a major public health concern as they are the leading cause of illness and death in many countries [Hanage, Schaffner, 2025]. The COVID-19 pandemic has been responsible for a significant economic and social burden and high mortality rates worldwide [Kobak, 2021; Wang et al., 2022]. Influenza also infects millions of people annually and causes hundreds of thousands of deaths [Li et al., 2019; Krammer et al., 2018]. In this regard, there is significant interest in the healthcare and academic communities in developing models and methods capable of modeling the spread of this infection in real time and predicting its spread within communities, with the aim of developing effective intervention strategies to control the outbreak and prevent its spread [Jefferson et al., 2023; Santangelo et al., 2023; Adly et al., 2020].

Agent-based models (ABMs) have become one of the most important tools in epidemiological research for simulating the spread of infectious respiratory diseases under various behavioral, demographic, and biological assumptions [Boyd et al., 2022]. Their ability to incorporate individual-level variability, reproduce complex interaction patterns, capture unexpected stochastic patterns, and simulate the effects of various interventions makes them particularly suitable for modeling epidemic dynamics in realistic populations [Tracy et al., 2018; Bonabeau, 2002; Ale et al., 2024]. Across the literature, two implementation principles for the ABMs dominate: (i) person – person contact networks with layered mixing (households, schools, workplaces, community), offering computational efficiency and clear levers for age- or group-specific policies; and (ii) person – location interactions that model mobility and venue occupancy, better capturing behavioral adaptation and place-based risks but demanding richer data and greater compute. Widely used toolkits exemplify these paradigms. Covasim provides an open-source platform for age-structured contact networks with detailed vaccination and immunity dynamics [Kerr et al., 2021], and has supported studies on vaccination prioritization and rollout optimization (e. g., Lombardy) [Cattaneo et al., 2022], dose-spacing and booster strategies [Moghadas et al., 2021; Pandey et al., 2023], and return-on-investment analyses [Sah et al., 2022]. The FRED platform anchors synthetic populations in census microdata and geo-linked places, enabling “digital twin” simulations of city/region dynamics, seasonality, and age-specific vaccine effects [Grefenstette et al., 2013; Krauland et al., 2022]. Geo-spatial ABMs have clarified venue- and neighborhood-level transmission patterns in large metropolitan areas [Aleta et al., 2022], while national and city-scale studies have integrated non-pharmaceutical interventions, testing, and reopening strategies [Chang et al., 2020; Scott et al., 2021]. A recent systematic review of 144 studies (2020–2024) systematizes these approaches and their limitations, highlighting the persistent trade-off between descriptive realism and computational tractability [Maslova et al., 2024].

Problem statement

Despite the power of ABMs, they face significant challenges in terms of computational complexity in their operation and calibration of epidemiological data [Zhang et al., 2020; Miksch et al., 2019]. Estimating parameters typically requires repeated simulations across large parameter spaces to determine plausible values for key epidemiological parameters [Sattenspiel et al., 2019], such as transmission rates and susceptibility ratios.

The high computational limitations of ABMs and the curse of high dimensions during calibration are a common challenge for a wide range of applications using ABMs across various fields and disciplines [Bergman et al., 2024; Norton et al., 2025]. Therefore, alleviating computational limitations and accelerating calibration is a broad and active area of research across various scientific disciplines [Cogno et al., 2024; Moya et al., 2021]. Some research has focused on simulation speed rather than the calibration process itself, reducing simulation execution time by improving computational efficiency through the use of parallel computing and hardware acceleration techniques to address long run-times [Perumalla, Aaby, 2008; Xiao et al., 2019]. Machine learning methods have also recently been used as surrogate models to direct simulation-based calibration, as they approximate the behavior of an ABM and enable rapid exploration of parameter-output relationships [Angione et al., 2022]. Other studies have focused on improving the calibration process itself, and several studies have explored meta-heuristic algorithms to improve calibration efficiency [Robertson et al., 2025; Vlad et al., 2024].

In epidemiology, epidemic data are sequential in nature and are calibrated with ABM output to recover underlying epidemiological parameters. Calibration and simulation-based inference are performed in real time and concurrently with the acquisition of observed epidemic data, which is computationally expensive. Surrogates can greatly facilitate this procedure by replacing or augmenting expensive simulators with fast, learnable approximations for calibration, forecasting, sensitivity analysis, and control. Classical statistical surrogates (e. g., Gaussian process or Kriging) have been applied to ABMs for emulation and uncertainty propagation, but can struggle in high dimensions or with nonsmooth responses [Norton et al., 2025]. Recent practice favors flexible machine learning surrogates — tree ensembles (XGBoost/Random Forest), support vector machines, neural networks, and, for spatial tasks, graph neural networks (GNNs) — trained on strategically sampled simulator runs [Angione et al., 2022; Perumal, van Zyl, 2025; Robertson et al., 2025]. Surrogate assisted optimization strategies such as Metric Stochastic Response Surfaces (MSRS) and DYCORS (dynamic coordinate search) guide sampling to promising regions of parameter space, achieving 2–4× speedups over baselines while matching or exceeding calibration accuracy; in head to head tests, MSRS+SVM best recovered ground truth parameters, whereas DYCORS+XGBoost gave the best fit speed trade off and $\sim 97\,\%$ similarity to real cumulative infections [Perumal, van Zyl, 2025]. Spatial surrogates built with GNNs can deliver county and age resolved forecasts for hundreds of regions in sub-second inference, naturally ingesting graph connectivity and contact pattern changes, though generalization to unseen graphs and weighted edges remains an open benchmark [Schmidt et al., 2024; Kraemer et al., 2025].

In this paper, we address the question: Can machine learning surrogates learn to relate epidemic trajectories (such as time series of incidence and prevalence) to underlying epidemiological parameters? That is, once trained, these models can quickly infer parameter estimates from observed epidemic data, reducing the time complexity of calibration. The proposed approach significantly reduces the calibration time by avoiding simulations over the entire space of possible epidemiological parameters. Experimental evaluation demonstrates the effectiveness of this approach on a synthetic epidemiological dataset generated through controlled ABM simulations.

The main contributions of this study can be summarized in framing the problem, applying three state-of-the-art models, and systematically comparing them. We formulate the task of calibrating ABMs against epidemic data as a supervised learning problem, where truncated sequences of epidemic trajectories (incidence and prevalence sequences) are mapped to underlying epidemiological parameters. To our knowledge, no study has mapped epidemic trajectories to underlying epidemiological parameters to reduce calibration time complexity. With this formulation, machine learning methods can be used to reduce calibration time complexity. We apply three models — Transformer Encoders (TEs), Bayesian Neural Networks (BNNs), and Mixture Density Networks (MDNs) — to the ABM calibration problem. We evaluated the performance of the models via accuracy, uncertainty estimation, and error structure.

Materials and methods

Data set

The simulation relies on an ABM framework [Leonenko et al., 2020] constructed around a detailed synthetic population. Synthetic population is a spatially constructed database of human agents representing the population of a city, region, or country. This database is comparable to a real population, but its records do not match real people, thus helping to avoid privacy issues. A synthetic population of St. Petersburg was used, representing virtual individuals and information about their households, workplaces, and schools. A pre-existing Python-based ABM framework implementing a Susceptible – Exposed – Infectious – Recovered (SEIR) structure was used to reproduce transmission dynamics within this synthetic population.

The selected ABM operates with several initial parameters including: the initial number of infectious individuals, the infection transmission coefficient lambda, which indicates how quickly an epidemic will spread, alpha, which determines fraction of non-immune individuals in general population, the seed parameter, which represents randomness in the model, and the number of days of tracking cases. The ABM simulates everyday interactions based on shared locations — households, workplaces, or schools. During the simulation, it calculates the probability of an infectious person coming into contact with a random susceptible person in the location, the probability of the susceptible person becoming infected, and finally, the number of infected individuals in the current location. The course of disease of an infected individual is modeled by a deterministic function which assumes a zero infectiousness during the first two days (incubation period) and a six-day infectious period with varying infectivity, after which the individual becomes recovered. The detailed description of synthetic population and ABM framework is discussed in the paper [Korzin et al., 2024]. The ABM's outputs are the incidence, which measures new cases of the disease in a population each day, and the prevalence, which measures the total number of cases present (new and old) each day.

The ABM was executed on the following input values: the synthetic data sample size is 500 000, the initial number of infectious individuals = 10, alpha within the range [0.05, 1] taken with a step of 0.05, lambda within the range [0.1, 1] taken with a step of 0.1, seed within the range [0, 9] taken with a step of 1, and the number of days of tracking cases = 80 days. To avoid any potential population differences, we adopted the incidence/prevalence rate per 10 000 population. Figure 1 shows the simulated data. The y-axis represents the daily incidence and prevalence rate, and the x-axis represents the daily time series. We note from the figure that the association of epidemic trajectories with epidemic coefficients is many to one, as different epidemic dynamics may be produced for similar coefficient configurations. This is due to the randomness achieved in the model through the seed parameter. We divided the data into two parts: 80 % of the data was the training set, and 20 % of the data was the test set.

Models

The problem can be formulated as: Let $X_{1:n}$ represent the first n days of the epidemic data series (incidence and prevalence) and let θ = (alpha, lambda). The learning task is to approximate $P(\theta \mid X_{1:n})$ using models that produce point estimates and confidence intervals. In this study, we evaluate three ML approaches that are well aligned with these challenges: TEs, BNNs, and MDNs. Each represents a distinct paradigm in contemporary ML: attention-based sequence modeling [Ahmed et al., 2023], probabilistic deep learning [Izmailov et al., 2021], and distributional regression [Bishop, 1994].

TEs have been widely adopted in sequence modeling because of their ability to capture long-term dependencies through self-concerning mechanisms, where representations of all time steps are computed simultaneously, enabling efficient modeling of long-term time dependencies [Vaswani et al., 2017]. This is advantageous for epidemic trajectories, in which early infection dynamics influence later

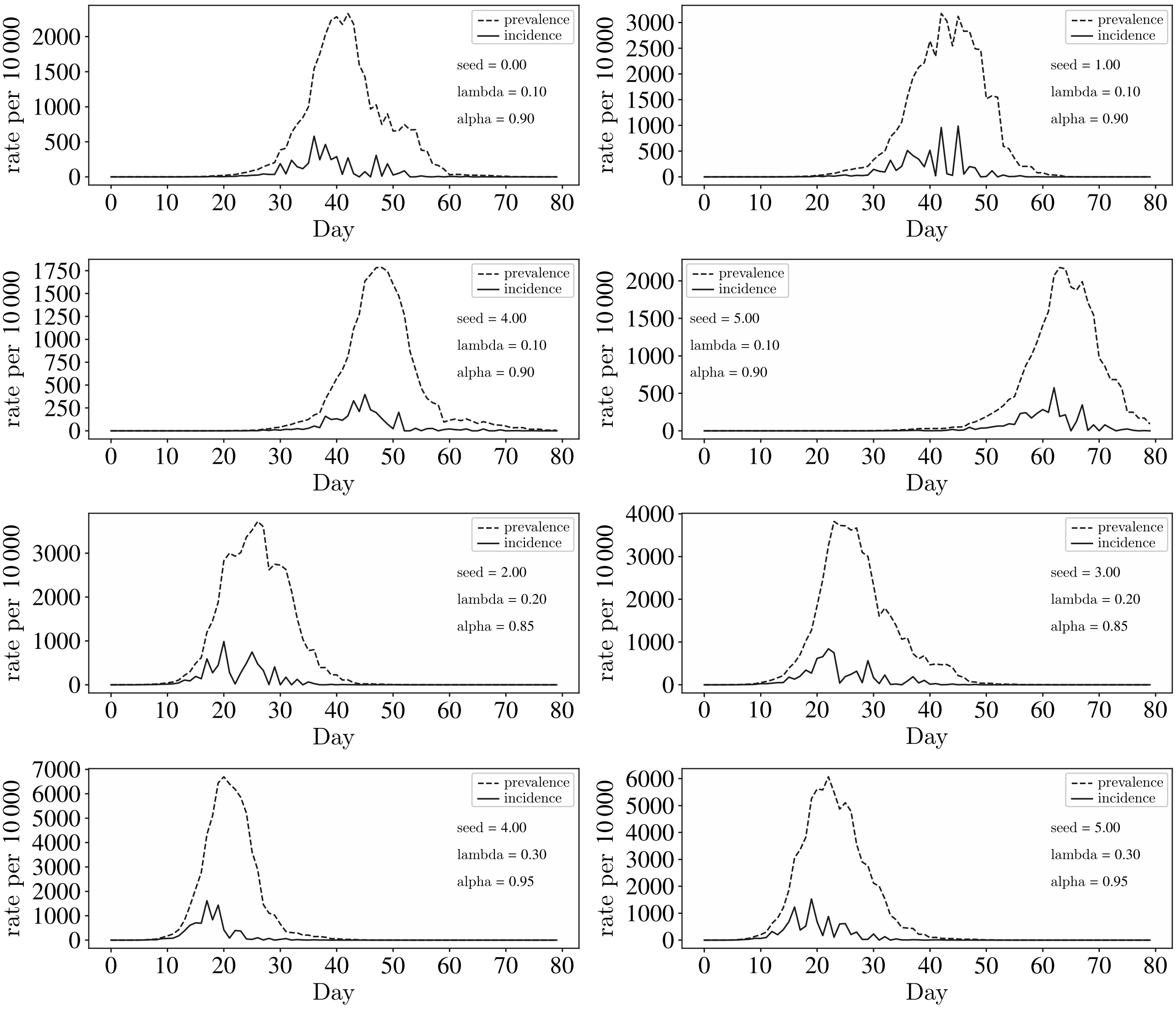

Figure 1. Examples of simulated data (prevalence and incidence rate per 10 000 over 80 days) generated by ABM framework for different inputs (seed, lambda, alpha)

prevalence patterns. TEs also support joint prediction of parameter values and associated uncertainty, making them suitable for inference under noisy epidemic observations.

MDNs perform distributional regression by representing the conditional output as a mixture of probability distributions [Bishop, 1994]. Instead of predicting a single deterministic estimate, MDNs output mixture weights, component means, and component variances, thereby modeling uncertainty and potentially multimodal parameter distributions. This is beneficial because epidemic data generated by ABMs can correspond to multiple plausible parameter combinations. By outputting these parameters, the network represents a probability density function, which can have multiple peaks, indicating multiple plausible outcomes for a given input. MDNs have been applied in inverse problems with multimodal solutions [Hepp et al., 2022], making them well-aligned with epidemiological calibration scenarios.

BNNs constitute a probabilistic framework in which neural-network weights are treated as random variables. Monte-Carlo dropout yields a tractable approximation to Bayesian inference by injecting stochasticity during both training and testing. This procedure estimates uncertainty, reflecting the model's confidence based on the data available [Gal, Ghahramani, 2016]. By applying the models to epidemic sequences and quantifying their effectiveness, this work contributes both methodological and practical insights into the suitability of ML approaches for ABM calibration. By systematically comparing these models, we aim to identify pathways toward more efficient and reliable calibration strategies that can enhance the role of ABMs in epidemic science and public health policy.

Metrics

We compared different models with different length of input using the mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination (R^2) as a measure of prediction accuracy. Comparing models in epidemiological parameter inference was not limited to accuracy alone, but was also accompanied by uncertainty measurement and error plots.

Results and discussion

In this study, the goal was to estimate the epidemiological parameters alpha and lambda from time series of incidence and prevalence. In the absence of surveillance data from real outbreaks, we used synthetic datasets generated from epidemiological simulations. The validity of using simulated data is based on the assumption that the simulation framework captures the underlying transmission dynamics. While no model can perfectly replicate reality, agent-based epidemic models have been shown to reproduce realistic outbreak dynamics across multiple diseases and contexts [Roche et al., 2011]. Therefore, inference methods that demonstrate robust performance on simulated datasets can be reasonably expected to maintain their effectiveness when applied to empirical data, provided the latter exhibits similar epidemiological characteristics.

Experiment design

The dataset we obtained contains all the various inputs from the ABM (alpha, lambda, seed) and the outputs of the ABM were represented as a sequence of incidence and prevalence over 80 days. In each experiment, one of the three models (TE, BNN, and MDN) was trained on a training set and tested on a test set. The inputs for each model were multivariate time series truncated from both incidence and prevalence sequences of length n, where n took one of the values (15, 20, 25, 30, 35, 40). During each experiment, the values of alpha and lambda were instantaneously assessed. Although all models were trained to infer both alpha and lambda parameters, the structure and interpretation of their predictions differed significantly.

In TE formulation, inputs are embedded and then mapped to a latent sequence representation using stacked encoder layers. This representation is subsequently aggregated and passed through two linear heads that produce a pair of vectors: one specifying mean for each target parameter and another specifying the associated standard deviations.

The MDN combines a recurrent backbone with a probabilistic mixture modeling head. In this setting, an LSTM-based encoder summarizes the inputs, and fully connected layers then return three sets of parameters: mixture weights, component means, and component scales. Together, these quantities define a full mixture distribution over the output parameters. The mean and variance of output parameters are calculated using weighted sums of the component parameters.

The BNN is constructed using fully connected layers. Its output is a single vector containing point predictions for the target parameters. Uncertainty is quantified indirectly: multiple stochastic forward passes produce a collection of parameter estimates, from which empirical means and standard deviations are computed. The model itself is deterministic in structure, but multiple stochastic forward passes yield empirical posterior samples. In contrast to the MDN or TE, the Bayesian network does not parameterize an explicit likelihood; instead, it samples from a distribution induced by dropout masks.

As a result, the output of each model is the mean and standard deviation for each output parameter. The mean was used as a point estimate and used with the standard deviation to calculate the 95 % confidence interval.

The total work included 18 experiments to train and test three models by changing the length of inputs. All models were implemented without hyper-parameter tuning, and all training and testing were performed on a personal computer. The metrics of all models for the parameter alpha are presented in Table 1, and those for the parameter lambda are presented in Table 2.

Table 1. Prediction errors of alpha across models and input lengths (n)

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.110	0.157	0.69	0.93	0.48	0.096	0.059	0.88	0.92	0.26
BNN	0.129	0.172	0.62	0.63	0.33	0.133	0.099	0.77	0.90	0.39
MDN	0.080	0.128	0.79	0.76	0.24	0.094	0.055	0.89	0.82	0.18
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.050	0.081	0.92	0.95	0.22	0.074	0.049	0.93	0.87	0.18
BNN	0.117	0.145	0.73	0.87	0.40	0.134	0.106	0.77	0.84	0.43
MDN	0.047	0.077	0.93	0.73	0.14	0.076	0.046	0.93	0.75	0.14
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	**0.047**	**0.068**	**0.94**	**0.94**	**0.20**	0.061	0.037	0.95	0.93	0.16
BNN	0.143	0.169	0.63	0.77	0.41	0.274	0.241	0.04	0.02	0.01
MDN	0.043	0.070	0.94	0.76	0.13	0.075	0.045	0.93	0.66	0.11

Table 2. Prediction errors of lambda across models and input lengths (n)

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.094	0.126	0.80	0.95	0.46	0.090	0.062	0.89	0.87	0.28
BNN	0.130	0.155	0.69	0.54	0.26	0.160	0.135	0.67	0.71	0.40
MDN	0.060	0.093	0.89	0.79	0.23	0.088	0.056	0.90	0.81	0.22
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.063	0.093	0.90	0.92	0.29	0.087	0.061	0.90	0.87	0.25
BNN	0.153	0.182	0.57	0.69	0.44	0.145	0.116	0.73	0.75	0.38
MDN	0.048	0.072	0.93	0.72	0.16	0.082	0.050	0.91	0.73	0.18
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.085	0.114	0.83	0.91	0.38	**0.087**	**0.063**	**0.90**	**0.93**	**0.31**
BNN	0.155	0.193	0.52	0.65	0.41	0.273	0.235	0.04	0.03	0.01
MDN	0.051	0.081	0.92	0.74	0.16	0.089	0.057	0.90	0.62	0.16

Overall performance

For the transmission parameter lambda, BNNs demonstrated moderate performance at shorter input lengths ($n = 15$–30) with R^2 values between 0.57 and 0.73. However, performance degraded significantly at longer horizons, particularly at $n = 40$, where R^2 collapsed to near zero ($R^2 = 0.04$) and the interval coverage vanished (coverage = 0.03, width = 0.01). This suggests that BNNs are highly sensitive to increasing sequence length. The posterior uncertainty estimates become unstable, reflecting either underfitting or oversmoothing effects when exposed to longer, noisier input sequences. The relatively wide confidence intervals indicate that the model compensates for prediction uncertainty by producing broad intervals, yet these do not consistently translate into high coverage. These results suggest that, while BNNs can provide reasonable estimates when temporal inputs are limited, they suffer from posterior instability and oversensitivity to increasing sequence length.

In contrast, TE provided more stable performance, maintaining high accuracy ($R^2 \geqslant 0.80$) and well-calibrated uncertainty across all n. Importantly, the predictive interval coverage remained relatively high and balanced (0.87–0.95) with widths (0.25–0.46) that were neither excessively wide nor overly narrow, though some variability is noted. For example, at $n = 20$ and $n = 30$, coverage dips slightly, indicating occasional underestimation of uncertainty. This indicates that the attention-based mechanism effectively captures temporal dependencies across varying horizons, enabling the model to maintain both accurate point estimates and well-calibrated uncertainty.

MDNs also performed strongly, achieving the lowest RMSE and MAE in several cases (notably $n = 25$ and 35), with consistently high R^2 values. However, their predictive intervals were consistently narrower than those of TEs, with widths typically between 0.16 and 0.23, leading to undercoverage (0.62–0.81). This pattern reflects a tendency of MDNs to underestimate epistemic uncertainty. While the point predictions are accurate, the model produces confidence intervals that are overly narrow, resulting in undercoverage. At $n = 40$, the performance remains stable in terms of RMSE, MAE and R^2, but the coverage drops to 0.62, highlighting this systematic limitation. This systematic underestimation of uncertainty, while advantageous for precise point prediction, limits their utility in scenarios where reliable uncertainty quantification is essential for decision-making.

A similar pattern is observed for the susceptible proportion parameter alpha in Table 1. TEs maintain robust accuracy and uncertainty calibration, MDNs provide highly accurate but less reliable uncertainty estimates, and BNNs demonstrate instability, particularly at longer input horizons.

These findings underscore that BNNs perform significantly worse compared to TEs and MDNs. While MDNs offer specific strengths in point accuracy, TEs provide the most consistent performance across metrics and horizons. This robustness is particularly advantageous in the context of epidemic model calibration, where both accurate parameter estimation and credible uncertainty quantification are crucial for informing public health decisions.

TE as a recommended model

All TE results starting from $n = 20$ are acceptable, since the accuracy is good and coverage is reasonable. The alpha prediction results shown in Table 1 demonstrate that the best result for input length $n = 35$ gives the following values of indicators: RMSE = 0.068, MAE = 0.047, $R^2 = 0.94$, coverage = 0.94, width = 0.20, which might be considered near-ideal (excellent accuracy with coverage ~ 0.95 at reasonable width). For $n = 25$ we have RMSE = 0.081, MAE = 0.05, $R^2 = 0.92$, coverage = 0.95, width = 0.22, it refers to a perfectly calibrated case with slightly worse accuracy than in the case of $n = 35$. The lambda prediction results shown in Table 2 demonstrate that the best results are obtained for input length $n = 40$ (RMSE = 0.087, MAE = 0.063, $R^2 = 0.90$, coverage = 0.93, width = 0.31), it refers to good accuracy, near-nominal coverage with moderate width. For $n = 30$ (RMSE = 0.087, MAE = 0.061, $R^2 = 0.90$, coverage = 0.87, width = 0.25), it refers to great accuracy, slight undercoverage. These results are consistent with the graphs in Fig. 2, which show the distribution of prediction errors for both alpha ($n = 35$) illustrated by Fig. 2, a and lambda ($n = 40$) illustrated by Fig. 2, *b*. The low bias of values is proved by a median being approximately equal zero, low variance is confirmed by a narrow inter-quartile range. A joint plot was also drawn to assess the alignment between the predicted and actual values in the test cases for both alpha ($n = 35$) shown in Fig. 2, *c* and lambda ($n = 40$) shown in Fig. 2, *d*. The alignment is demonstrated by points lying close to the diagonal ($y = x$) and the overlapping of marginal histograms.

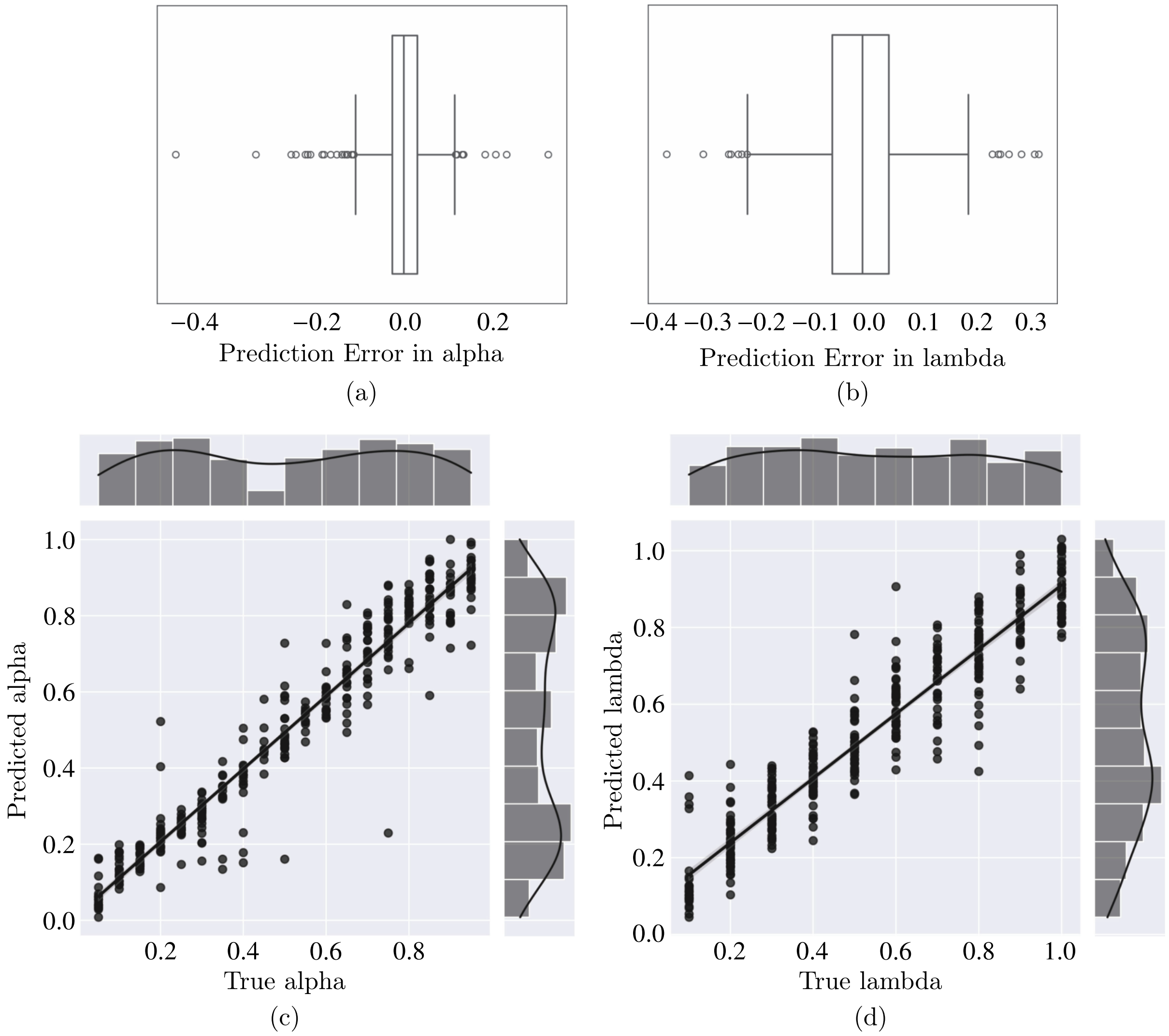

Figure 2. Distribution of prediction errors and alignment between predicted and true values across test cases: (a) boxplot for prediction errors in alpha using the TE model at $n = 35$; (b) boxplot for prediction errors in lambda using the TE model at $n = 40$; (c) jointplot for predicted vs. true alpha using the TE model at $n = 35$; (d) jointplot for predicted vs. true lambda using the TE model at $n = 40$

The case of limited input

In real epidemic surveillance, it is not guaranteed that the statistics both on disease incidence and prevalence is available as modeling input. Particularly, for the case of acute respiratory infections, the COVID-19 statistics we employ in modeling contains both incidence and prevalence, whereas for influenza only incidence is available.

To assess how the surrogate approaches will perform in that situation, two additional sets of experiments were conducted with new input configurations: prevalence-only time series (see Tables 3, 4) and incidence-only time series (see Tables 5, 6).

For alpha inference, prevalence alone was highly informative, yielding performance close to or comparable with the incidence+prevalence input configuration. This is expected because alpha directly constrains the size of the group of susceptible individuals and therefore the overall epidemic size, both of which are reflected more clearly in cumulative metrics. In contrast, lambda inference was more challenging. This was reflected in lower R^2 values for TE at intermediate input lengths and in severe

Table 3. Prediction errors of alpha across models and input lengths (*n*) for prevalence-only input

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.085	0.128	0.79	0.94	0.4	0.071	0.111	0.84	0.96	0.37
BNN	0.115	0.149	0.72	0.68	0.31	0.242	0.275	0.03	0.02	0.02
MDN	0.076	0.129	0.79	0.88	0.27	0.053	0.095	0.88	0.89	0.19
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.053	0.092	0.89	0.95	0.27	0.046	0.081	0.92	0.94	0.2
BNN	0.093	0.14	0.75	0.88	0.45	0.123	0.161	0.67	0.86	0.4
MDN	0.052	0.084	0.91	0.85	0.19	0.053	0.087	0.9	0.75	0.16
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.041	0.068	0.94	0.93	0.18	0.045	0.07	0.94	0.94	0.23
BNN	0.243	0.277	0.02	0.01	0.03	0.12	0.171	0.63	0.81	0.43
MDN	0.041	0.068	0.94	0.77	0.13	0.05	0.086	0.91	0.66	0.12

Table 4. Prediction errors of lambda across models and input lengths (*n*) for prevalence-only input

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.078	0.1026	0.86	0.96	0.4	0.084	0.112	0.84	0.95	0.42
BNN	0.105	0.128	0.79	0.72	0.29	0.231	0.27	0.06	0.04	0.02
MDN	0.064	0.099	0.87	0.83	0.27	0.048	0.074	0.93	0.84	0.22
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.095	0.126	0.8	0.97	0.52	0.104	0.139	0.75	0.94	0.49
BNN	0.11	0.14	0.75	0.76	0.4	0.134	0.167	0.64	0.81	0.41
MDN	0.047	0.08	0.92	0.83	0.21	0.052	0.083	0.91	0.75	0.18
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.096	0.127	0.79	0.95	0.47	0.105	0.138	0.75	0.95	0.56
BNN	0.235	0.274	0.03	0.01	0.03	0.153	0.18	0.58	0.68	0.45
MDN	0.05	0.084	0.91	0.68	0.15	0.058	0.093	0.89	0.7	0.18

instability for BNNs, which failed to produce reliable posterior estimates at $n = 20$ and $n = 35$. MDNs again delivered accurate point predictions but narrower-than-optimal intervals. These results indicate that prevalence alone provides sufficient information for alpha but is less reliable for lambda.

Incidence alone provides only limited information about alpha. No improvement was shown except in the BNN model. TE shows reduced accuracy and less stable uncertainty when relying solely on incidence, while MDN gave more accurate point predictions, but at the expense of coverage. Incidence reflects new infections and is therefore sensitive to lambda, but the sensitivity is not sufficient for reliable inference when incidence is the only signal available. Across TE and MDN models, incidence-only inputs produce noticeably higher errors and wider uncertainty intervals than joint inputs.

Limitations

All performance measures are reported as point estimates based on a single data split. This approach does not account for variability that may arise from different data splits or random initialization. More rigorous validation requires interval assessment and the construction of confidence

Table 5. Prediction errors of alpha across models and input lengths (n) for incidence-only input

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.114	0.168	0.64	0.95	0.53	0.093	0.144	0.74	0.95	0.42
BNN	0.101	0.146	0.73	0.76	0.3	0.084	0.116	0.83	0.87	0.34
MDN	0.076	0.131	0.78	0.69	0.22	0.053	0.097	0.88	0.68	0.16
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.07	0.11	0.84	0.9	0.3	0.067	0.113	0.84	0.94	0.33
BNN	0.077	0.106	0.86	0.91	0.33	0.064	0.102	0.87	0.89	0.32
MDN	0.05	0.091	0.89	0.61	0.13	0.048	0.091	0.89	0.52	0.09
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.06	0.105	0.86	0.94	0.27	0.05	0.09	0.89	0.91	0.21
BNN	0.056	0.077	0.92	0.89	0.26	0.05	0.069	0.94	0.96	0.31
MDN	0.037	0.064	0.95	0.55	0.07	0.034	0.061	0.95	0.63	0.07

Table 6. Prediction errors of lambda across models and input lengths (n) for incidence-only input

	$n = 15$					$n = 20$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.116	0.162	0.66	0.92	0.56	0.131	0.171	0.62	0.96	0.69
BNN	0.087	0.117	0.82	0.78	0.28	0.096	0.127	0.79	0.75	0.29
MDN	0.065	0.108	0.85	0.68	0.22	0.067	0.106	0.85	0.67	0.19
	$n = 25$					$n = 30$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.154	0.2	0.48	0.93	0.66	0.165	0.208	0.44	0.93	0.71
BNN	0.083	0.117	0.82	0.85	0.32	0.087	0.119	0.82	0.83	0.32
MDN	0.075	0.118	0.82	0.49	0.15	0.074	0.117	0.82	0.38	0.11
	$n = 35$					$n = 40$				
MODEL	MAE	RMSE	R^2	Coverage	Width	MAE	RMSE	R^2	Coverage	Width
TE	0.126	0.17	0.62	0.92	0.6	0.109	0.147	0.72	0.89	0.47
BNN	0.096	0.124	0.80	0.75	0.27	0.089	0.119	0.82	0.84	0.33
MDN	0.06	0.091	0.89	0.46	0.12	0.061	0.093	0.89	0.56	0.14

intervals for each measure. This approach would provide stronger statistical evidence regarding the comparative performance of the models and reduce the possibility that observed differences are due to specific experimental conditions.

Another open question is the following: are the demonstrated results on surrogate performance generalizable enough? Particularly, can we claim that the models under consideration show comparable results on more complex agent-based models with a larger number of parameters? In the absence of empirical evidence from relevant works, one can rely on the theoretical scalability of models used. Theoretically, attention-based architectures tend to exhibit good scalability and likely to remain competitive because their representational capacity increases with model depth and width. Mixture-density also remains tractable when the dimensionality of the output distribution is moderate. BNNs generally scale less efficiently, so it might experience problems on more complex ABM architectures.

Conclusion

In this work, we applied and evaluated three machine learning models with different methodologies to alleviate the computational limitations in the inverse problem of calibrating an

ABM simulating the spread of respiratory infection in Saint Petersburg. Under the assumption that the simulation framework captures the underlying transmission dynamics, we conclude that machine learning-based models are able to alleviate computational limitations and can be used in a practical and straightforward manner. We plan to address interval assessment in future work, particularly in the context of real epidemiological data, where interval estimates will be critical for drawing robust conclusions.

References

Adly A. S., Adly A. S., Adly M. S. Approaches based on artificial intelligence and the Internet of Intelligent Things to prevent the spread of COVID-19: Scoping review // Journal of Medical Internet Research. — 2020. — Vol. 22, No. 8. — P. e19104. — https://doi.org/10.2196/19104

Ahmed S., Nielsen I. E., Tripathi A., Siddiqui S., Ramachandran R. P., Rasool G. Transformers in time-series analysis: A tutorial // Circuits, Systems, and Signal Processing. — 2023. — Vol. 42, No. 12. — P. 7433–7466. — https://doi.org/10.1007/s00034-023-02454-8

Ale S., Hunter E., Kelleher J. D. Agent based modelling of blood borne viruses: a scoping review // BMC Infectious Diseases. — 2024. — Vol. 24, No. 1. — P. 1411. — https://doi.org/10.1186/s12879-024-10271-w

Aleta A., Martín-Corral D., Bakker M. A., Pastore y Piontti A., Ajelli M., Litvinova M., Chinazzi M., Dean N. E., Halloran M. E., Longini I. M., Pentland A., Vespignani A., Moreno Y., Moro E. Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas // Proceedings of the National Academy of Sciences. — 2022. — Vol. 119, No. 26. — P. e2112182119. — DOI: 10.1073/pnas.2112182119

Angione C., Silverman E., Yaneske E. Using machine learning as a surrogate model for agent-based simulations // PLoS One. — 2022. — Vol. 17, No. 2. — P. e0263150. — https://doi.org/10.1371/journal.pone.0263150

Bergman D. R., Norton K. A., Jain H. V., Jackson T. Connecting agent-based models with high-dimensional parameter spaces to multidimensional data using SMoRe ParS: A surrogate modeling approach // Bulletin of Mathematical Biology. — 2024. — Vol. 86, No. 1. — P. 11. — https://doi.org/10.1007/s11538-023-01240-6

Bishop C. M. Mixture density networks // Technical Report NCRG/94/004. — Neural Computing Research Group, Aston University, 1994.

Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems // Proceedings of the National Academy of Sciences of the USA. — 2002. — Vol. 99. — P. 7280–7287. — https://doi.org/10.1073/pnas.082080899

Boyd J., Wilson R., Elsenbroich C., Heppenstall A., Meier P. Agent-based modelling of health inequalities following the complexity turn in public health: a systematic review // International Journal of Environmental Research and Public Health. — 2022. — Vol. 19, No. 24. — P. 16807. — https://doi.org/10.3390/ijerph192416807

Cattaneo A., Vitali A., Mazzoleni M., Previdi F. An agent-based model to assess large-scale COVID-19 vaccination campaigns for the Italian territory: The case study of Lombardy region // Computer Methods and Programs in Biomedicine. — 2022. — Vol. 224. — P. 107029. — DOI: 10.1016/j.cmpb.2022.107029

Chang S. L., Harding N., Zachreson C., Cliff O. M., Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia // Nature Communications. — 2020. — Vol. 11. — P. 5710. — DOI: 10.1038/s41467-020-19393-6

Cogno N., Axenie C., Bauer R., Vavourakis V. Agent-based modeling in cancer biomedicine: applications and tools for calibration and validation // Cancer Biology & Therapy. — 2024. — Vol. 25, No. 1. — P. 2344600. — https://doi.org/10.1080/15384047.2024.2344600

Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning // Proceedings of the International Conference on Machine Learning. — 2016. — P. 1050–1059.

Grefenstette J. J., Brown S. T., Rosenfeld R., DePasse J., Stone N. T., Cooley P. C., Wheaton W. D., Fyshe A., Galloway D. D., Sriram A., Guclu H., Abraham T., Burke D. S. FRED (a Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations // BMC Public Health. — 2013. — Vol. 13. — P. 940. — DOI: 10.1186/1471-2458-13-940

Hanage W. P., Schaffner W. Burden of acute respiratory infections caused by influenza virus, respiratory syncytial virus, and SARS-CoV-2 with consideration of older adults: A narrative review // Infectious Diseases and Therapy. — 2025. — Vol. 14, Suppl. 1. — P. 5–37. — https://doi.org/10.1007/s40121-024-01080-4

Hepp T., Zierk J., Rauh M., Metzler M., Seitz S. Mixture density networks for the indirect estimation of reference intervals // BMC Bioinformatics. — 2022. — Vol. 23, No. 1. — P. 307. — https://doi.org/10.1186/s12859-022-04846-0

Izmailov P., Vikram S., Hoffman M. D., Wilson A. G. G. What are Bayesian neural network posteriors really like? // International Conference on Machine Learning. — PMLR, 2021. — P. 4629–4640.

Jefferson T., Dooley L., Ferroni E., Al-Ansary L. A., van Driel M. L., Bawazeer G. A., Jones M. A., Hoffmann T. C., Clark J., Beller E. M., Glasziou P. P. Physical interventions to interrupt or reduce the spread of respiratory viruses // Cochrane Database of Systematic Reviews. — 2023. — No. 1. — https://doi.org/10.1002/14651858.CD006207.pub6

Kerr C. C., Stuart R. M., Mistry D., Abeysuriya R. G., Rosenfeld K., Hart G. R., Núñez R. C., Cohen J. A., Selvaraj P., Hagedorn B., George L., Jastrzębski M., Izzo A. S., Fowler G., Palmer A., Delport D., Scott N., Kelly S. L., Bennette C. S., Wagner B. G., Chang S. T., Oron A. P., Wenger E. A., Panovska-Griffiths J., Famulare M., Klein D. J. Covasim: an agent-based model of COVID-19 dynamics and interventions // PLoS Computational Biology. — 2021. — Vol. 17, No. 7. — P. e1009149. — DOI: 10.1371/journal.pcbi.1009149

Kobak D. Excess mortality reveals Covid's true toll in Russia // Significance (Oxford, England). — 2021. — Vol. 18, No. 1. — P. 16. — https://doi.org/10.1111/1740-9713.01486

Korzin A. I., Kaparulin T. I., Leonenko V. N. Assessing the applicability of the multiagent modeling approach to the epidemic surveillance of COVID-19 in Russian cities // 2024 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), Novosibirsk, Russian Federation. — 2024. — P. 237–242. — https://doi.org/10.1109/SIBIRCON63777.2024.10758442

Kraemer M. U. G., Tsui J. L.-H., Chang S. Y., Lytras S., Khurana M. P., Vanderslott S., Bajaj S., Scheidwasser N., Curran-Sebastian J. L., Semenova E., Zhang M., Unwin H. J. T., Watson O. J., Mills C., Dasgupta A., Ferretti L., Scarpino S. V., Koua E., Morgan O., Tegally H., Paquet U., Moutsianas L., Fraser C., Ferguson N. M., Topol E. J., Duchêne D. A., Stadler T., Kingori P., Parker M. J., Dominici F., Shadbolt N., Suchard M. A., Ratmann O., Flaxman S., Holmes E. C., Gomez-Rodriguez M., Schölkopf B., Donnelly C. A., Pybus O. G., Cauchemez S., Bhatt S. Artificial intelligence for modelling infectious disease epidemics // Nature. — 2025. — Vol. 638. — P. 623–635. — DOI: 10.1038/s41586-024-08564-w

Krammer F., Smith G. J. D., Fouchier R. A. M., Peiris M., Kedzierska K., Doherty P. C., Palese P., Shaw M. L., Treanor J., Webster R. G., García-Sastre A. Influenza // Nature Reviews Disease Primers. — 2018. — Vol. 4, No. 1. — P. 3. — https://doi.org/10.1038/s41572-018-0002-y

Krauland M. G., Galloway D. D., Raviotta J. M., Zimmerman R. K., Roberts M. S. Impact of low rates of influenza on next-season influenza infections // American Journal of Preventive Medicine. — 2022. — Vol. 62, No. 4. — P. 503–510. — DOI: 10.1016/j.amepre.2021.11.007

Leonenko V., Arzamastsev S., Bobashev G. Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling // Journal of Computational Science. — 2020. — Vol. 44. — P. 101156.

Li L., Liu Y., Wu P., Peng Z., Wang X., Chen T., Wong J. Y. T., Yang J., Bond H. S., Wang L., Lau Y. C., Zheng J., Feng S., Qin Y., Fang V. J., Jiang H., Lau E. H. Y., Liu S., Qi J., Zhang J., Yang J., He Y., Zhou M., Cowling B. J., Feng L., Yu H. Influenza-associated excess respiratory mortality in China, 2010–15: a population-based study // The Lancet Public Health. — 2019. — Vol. 4, No. 9. — P. e473–e481. — https://doi.org/10.1016/S2468-2667(19)30163-X

Maslova I. I., Manolov A. I., Glushchenko O. E., Kozlov I. E., Tsurkis V. I., Popov N. S., Samoilov A. E., Lukashev A. N., Ilina E. N. Limitations in creating artificial populations in agent-based epidemic modeling: a systematic review // Journal of Microbiology, Epidemiology and Immunobiology. — 2024. — Vol. 101, No. 4. — P. 530–545. — DOI: 10.36233/0372-9311-467

Miksch F., Jahn B., Espinosa K. J., Chhatwal J., Siebert U., Popper N. Why should we apply ABM for decision analysis for infectious diseases? — An example for dengue interventions // PLoS One. — 2019. — Vol. 14, No. 8. — P. e0221564. — https://doi.org/10.1371/journal.pone.0221564

Moghadas S. M., Vilches T. N., Zhang K., Nourbakhsh S., Sah P., Fitzpatrick M. C., Galvani A. P. Evaluation of COVID-19 vaccination strategies with a delayed second dose // PLoS Biology. — 2021. — Vol. 19, No. 4. — P. e3001211. — DOI: 10.1371/journal.pbio.3001211

Moya I., Chica M., Cordon O. Evolutionary multiobjective optimization for automatic agent-based model calibration: A comparative study // IEEE Access. — 2021. — Vol. 9. — P. 55284–55299. — https://doi.org/10.1109/ACCESS.2021.3070071

Norton K. A., Bergman D., Jain H. V., Jackson T. Advances in surrogate modeling for biological agent-based simulations: Trends, challenges, and future prospects // arXiv preprint. — 2025. — arXiv:2504.11617

Pandey A., Fitzpatrick M. C., Moghadas S. M., Vilches T. N., Ko C., Vasan A., Galvani A. P. Modelling the impact of a high-uptake bivalent booster scenario on the COVID-19 burden and healthcare costs in New York City // The Lancet Regional Health – Americas. — 2023. — Vol. 24. — P. 100555. — DOI: 10.1016/j.lana.2023.100555

Perumal R., van Zyl T. L. Surrogate-assisted strategies: the parameterisation of an infectious disease agent-based model // Neural Computing and Applications. — 2025. — Vol. 37. — P. 627–638. — DOI: 10.100/s00521-022-07476-y

Perumalla K. S., Aaby B. G. Data parallel execution challenges and runtime performance of agent simulations on GPUs // Proceedings of the 2008 Spring Simulation Multiconference. — 2008. — P. 116–123.

Robertson C., Safta C., Collier N., Ozik J., Ray J. Bayesian calibration of stochastic agent-based model via random forest // Statistics in Medicine. — 2025. — Vol. 44, No. 6. — P. e70029. — https://doi.org/10.1002/sim.70029

Roche B., Drake J. M., Rohani P. An agent-based model to study the epidemiological and evolutionary dynamics of influenza viruses // BMC Bioinformatics. — 2011. — Vol. 12, No. 1. — P. 87. — https://doi.org/10.1186/1471-2105-12-87

Sah P., Vilches T. N., Moghadas S. M., Pandey A., Gondi S., Schneider E. C., Singer J., Chokshi D. A., Galvani A. P. Return on investment of the COVID-19 vaccination campaign in New York City // JAMA Network Open. — 2022. — Vol. 5, No. 11. — P. e2243127. — DOI: 10.1001/jamanetworkopen.2022.43127

Santangelo O. E., Gentile V., Pizzo S., Giordano D., Cedrone F. Machine learning and prediction of infectious diseases: a systematic review // Machine Learning and Knowledge Extraction. — 2023. — Vol. 5, No. 1. — P. 175–198. — https://doi.org/10.3390/make5010013

Sattenspiel L., Dimka J., Orbann C. Using cultural, historical, and epidemiological data to inform, calibrate, and verify model structures in agent-based simulations // Mathematical Biosciences and Engineering. — 2019. — Vol. 16, No. 4. — P. 3071–3093. — https://doi.org/10.3934/mbe.2019152

Schmidt A., Zunker H., Heinlein A., Kuhn M. J. Graph neural network surrogates to leverage mechanistic expert knowledge towards reliable and immediate pandemic response // arXiv preprint. — 2024. — arXiv:2411.06500

Scott N., Palmer A., Delport D., Abeysuriya R., Stuart R. M., Kerr C. C., Mistry D., Klein D. J., Sacks-Davis R., Heath K., Hainsworth S. W., Pedrana A., Stoove M., Wilson D., Hellard M. E. Modelling the impact of relaxing COVID-19 control measures during a period of low viral transmission // Medical Journal of Australia. — 2021. — Vol. 214, No. 2. — P. 79–83. — DOI: 10.5694/mja2.50845

Tracy M., Cerdá M., Keyes K. M. Agent-based modeling in public health: current applications and future directions // Annual Review of Public Health. — 2018. — Vol. 39, No. 1. — P. 77–94. — https://doi.org/10.1146/annurev-publhealth-040617-014317

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems. — 2017. — Vol. 30.

Vlad A. I., Romanyukha A. A., Sannikova T. E. Parameter tuning of agent-based models: Metaheuristic algorithms // Mathematics. — 2024. — Vol. 12, No. 14. — P. 2208. — https://doi.org/10.3390/math12142208

Wang H., Paulson K. R., Pease S. A., Watson S., Comfort H., Zheng P., Aravkin A. Y., Bisignano C., Barber R. M., Alam T., Fuller J. E. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21 // The Lancet. — 2022. — Vol. 399, No. 10334. — P. 1513–1536. — https://doi.org/10.1016/S0140-6736(21)02796-3

Xiao J., Andelfinger P., Eckhoff D., Cai W., Knoll A. A survey on agent-based simulation using hardware accelerators // ACM Computing Surveys (CSUR). — 2019. — Vol. 51, No. 6. — P. 1–35. — https://doi.org/10.1145/3291048

Zhang Y., Li Z., Zhang Y. Validation and calibration of an agent-based model: A surrogate approach // Discrete Dynamics in Nature and Society. — 2020. — Vol. 2020, No. 1. — P. 6946370. — https://doi.org/10.1155/2020/6946370